Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently amended) A method for constructing a variant set for modifying a biopolymer of interest, the method comprising:

a) identifying, ~~using a plurality of rules,~~ a plurality of positions in said biopolymer of interest and, for each respective position in said plurality of positions, one or more substitutions for the respective position, wherein the plurality of positions and the one or more substitutions for each respective position in the plurality of positions collectively define a biopolymer sequence space; and

b) selecting a first plurality of variants of the biopolymer of interest thereby forming a variant set, wherein said variant set comprises ~~a plurality of variants of said biopolymer of interest and wherein said variant set is~~ a subset of said biopolymer sequence space;

c) measuring a property of all or a portion of the variants in the variant set; and

d) modeling a sequence-activity relationship between (i) one or more substitutions at one or more positions of the biopolymer of interest represented by the variant set and (ii) the property measured for all or the portion of the variants in the variant set and deriving from said sequence-activity relationship (i) a value for the contribution to the measured property by the one or more substitutions at one or more positions of the biopolymer of interest, and (ii) a value quantifying the confidence with which the contribution to the measured property by the one or more substitutions at one or more positions of the biopolymer of interest can be assigned[[;]] ~~and~~

~~e) redefining the variant set to comprise variants that include substitutions in the plurality of positions that are selected based on a function of the sequence-activity relationship~~.

2-116. (Cancelled)

117. (New) The method of claim 1, the method further comprising:

e) defining a new variant set for the biopolymer of interest that comprises variants that include substitutions in the plurality of positions that are selected based on a function of the sequence-activity relationship.

118. (New) The method of claim 117, the method further comprising:

f) measuring a property of all or a portion of the variants in the new variant set.

119. (New) The method of claim 1, wherein the plurality of positions and the one or more substitutions for each respective position in the plurality of positions are identified using a plurality of rules.

120. (New) The method of claim 119, wherein the plurality of rules comprises two or more rules selected from the group consisting of:

(i) the favorability of a substitution calculated from a substitution matrix;

(ii) the probability of a substitution calculated from a conservation index;

(iii) the proximity of a position to a structurally defined region within the biopolymer,

(iv) the presence of a substitution in a homologous biopolymer;

(v) the favorability of a substitution calculated from a comparison of homologous sequences;

(vi) the mutability of a position calculated from a comparison of homologous sequences;

(vii) the favorability of a substitution calculated from a comparison of homologous structures; and

(viii) the mutability of a position calculated from a comparison of homologous structures.

121. (New) The method of claim 1, wherein the variant set is selected using at least one selection criterion that results in enrichment for pairwise uniqueness of substitutions at positions in the plurality of positions,

122. (New) The method of claim 1, wherein the variant set consists of fewer than 1000 variants.

123. (New) The method of claim 1, wherein the variant set consists of fewer than 250 variants.

124. (New) The method of claim 1, wherein the variant set consists of fewer than 100 variants.

125. (New) The method of claim 1, wherein variants in the variant set contain fewer than 5 substitutions.

126. (New) The method of claim 117, wherein the new variant set comprises variants of the biopolymer that have one or more substitutions at one or more positions that are not encompassed by the biopolymer sequence space of step a).

127. (New) The method of claim 121, wherein the at least one selection criterion is a coverage algorithm that distributes the substitutions in the plurality of positions across variants in the variant set.

128. (New) The method of claim 1, wherein the value quantifying the confidence with which the contribution to the measured property of the one or more substitutions at one or more positions of the biopolymer of interest can be assigned is a standard deviation of the value for the contribution to the measured property of the one or more substitutions at one or more positions of the biopolymer of interest.

129. (New) The method of claim 117, wherein variants in the new variant set differ by fewer than 5 substitutions from at least one biopolymer for which the property has already been measured.

130. (New) The method of claim 117, wherein the defining of a new variant set step e) further comprises:

computing a contribution score quantifying the contribution to the property of each one or more substitutions at the plurality of positions using the sequence-activity relationship;

computing a confidence score quantifying the confidence with which the contribution score is set;

computing a modified contribution score by modifying the contribution score based on a function of the confidence score; and

wherein said function of said sequence-activity relationship comprises using the modified contribution score for each one or more substitutions at the plurality of positions as a basis for including or excluding substitutions from the new variant set.

131. (New) The method of claim 130, the method further comprising:

ranking the one or more substitutions at the plurality of positions based on the modified contribution score; and

wherein said function of said sequence-activity relationship comprises accepting a predetermined percentage of the top ranked substitutions in the one or more substitutions at the plurality of positions for inclusion in the new variant set.

132. (New) The method of claim 130, wherein a respective one or more substitutions at the plurality of positions is selected for inclusion in the new variant set when the modified contribution score exceeds a predetermined value.

133. (New) The method of claim 118, the method further comprising repeating steps b) through f), until a variant in the new variant set exhibits a value for the property that exceeds a predetermined value.

134. (New) The method of claim 133, wherein the predetermined value is a value that is greater than the value for the property that is exhibited by the biopolymer of interest.

135. (New) The method of claim 118, the method further comprising repeating steps b) though f), until a variant in the variant set exhibits a value for the property that is less than a predetermined value.

136. (New) The method of claim 135, wherein the predetermined value is a value that is less than the value for the property that is exhibited by the biopolymer of interest.

137. (New) The method of claim 1, wherein the sequence-activity relationship comprises a plurality of values and wherein each value in the plurality of values describes a relationship between the property and:

(i) a substitution at a position in the plurality of positions represented by the all or the portion of the variants in the variant set,

(ii) a plurality of substitutions at a position in the plurality of positions represented by the all or the portion of the variants in the variant set, or

(iii) one or more substitutions in one or more positions in the plurality of positions represented by the all or the portion of the variants in the variant set.

138. (New) The method of claim 137, wherein the modeling comprises regressing:

$$V_{measured} = W_{11}P_1S_1 + W_{12}P_1S_2 + \ldots + W_{1N}P_1S_N + \ldots + W_{M1}P_MS_1 + W_{M2}P_MS_2 + \ldots + W_{MN}P_MS_N$$

wherein,

$V_{measured}$ represents the property measured in variants in the variant set;

W_{MN} = is a value in the plurality of values;

P_M = is a position in the biopolymer of interest in the plurality of positions in the biopolymer of interest; and

S_N = is a substitution in the one or more positions for a position in the plurality of positions in the biopolymer of interest.

139. (New) The method of claim 138, wherein the regressing comprises linear regression, non-linear regression, logistic regression, multivariate data analysis, or partial least squares projection to latent variables.

140. (New) The method of claim 1, wherein the modeling step d) comprises computation of a neural network, computation of a Bayesian model, a generalized additive model, a support vector machine, machine learning, or classification using a regression tree.

141. (New) The method of claim 1, wherein the modeling step d) comprises boosting or adaptive boosting.

142. (New) The method of claim 117, wherein the defining of a new variant set step e) comprises:

computing a contribution score quantifying the contribution to the property of each of one or more substitutions at the plurality of positions using the sequence-activity relationship;

computing a confidence score quantifying the confidence with which the contribution score is set;

computing a modified contribution score by modifying the contribution score based on a function of the confidence score; and

computing a predicted score for a population of variants of the biopolymer of interest using the modified contribution score to predict the contribution of one or more substitutions at one or more positions in the plurality of positions in the biopolymer of interest, wherein each variant in the population of variants includes a substitution at one or more positions in the plurality of positions in the biopolymer of interest; and

selecting the new variant set from among the population of variants based on the predicted score received by each variant in the set of variants.

143. (New) The method of claim 142, the method further comprising:

ranking the population of variants, wherein each variant in the population of variants is ranked based on the predicted score received by the variant based upon the sequence-activity relationship; and

the selecting comprising accepting a predetermined percentage of the top ranked variants in the population of variants for the variant set.

144. (New) The method of claim 142, wherein a respective variant in the population of variants is selected for the new variant set when the predicted score of the respective variant exceeds a predetermined value.

145. (New) The method of claim 117, wherein the defining of a new variant set step e) further comprises defining the new variant set to comprise one or more variants each having a substitution in a position in the plurality of positions not present in any variant in the variant set selected by the selecting step b).

146. (New) The method of claim 145, wherein the defining of a new variant set step e) further comprises redefining the variant set to comprise one or more variants each having a substitution in a position in the plurality of positions not present in any variant in the variant set selected by the selecting step (b).

147. (New) The method of claim 117, wherein the plurality of positions and the one or more substitutions for each respective position in the plurality of positions are identified using a plurality of rules; and wherein

the contribution of each respective rule in the plurality of rules to the biopolymer sequence space is independently weighted by a rule weight in a plurality of rule weights corresponding to the respective rule; and

the method further comprises, prior to the defining of a new variant set step e), the steps of:

adjusting one or more rule weights in the plurality of rule weights based on a comparison, for each respective substitution at each position in the plurality of positions in the variant set, (i) a value derived for the respective substitution at each position in the plurality of positions from the sequence-activity relationship, and (ii) a score assigned by the plurality of rules to the respective substitution at each position in the plurality of positions; and

repeating the identifying step using the rule weights, thereby redefining the plurality of positions and, for each respective position in the plurality of positions, redefining the one or more substitutions for the respective position; and wherein

the defining of a new variant set step e) further comprises redefining the variant set to comprise one or more variants each having a substitution in a position in the redefined plurality of positions not present in any variant in the variant set selected by the initial selecting step b).

148. (New) The method of claim 117 wherein

the modeling a sequence-activity relationship d) further comprises modeling a plurality of sequence-activity relationships, wherein each respective sequence-activity relationship in the plurality of sequence-activity relationships describes the relationship between (i) one or more substitutions at one or more positions of the biopolymer of interest represented by the variant set and (ii) the property measured for all or the portion of the variants in the variant set; and

the defining the variant set e) comprises redefining the variant set to comprise variants that include substitutions in the plurality of positions that are selected based on a combination function of the plurality of sequence-activity relationships.

149. (New) The method of claim 148, the method further comprising:

measuring a property of all or a portion of the variants in the new variant set; and

weighting each respective sequence-activity relationship in the plurality of sequence activity relationships based on an agreement between (i) measured values for the property of variants in the new variant set and (ii) values for the property of variants in the new variant set that were predicted by the respective sequence-activity relationship, wherein

a first sequence-activity relationship that achieves better agreement between measured and predicted values than a second sequence-activity relationship receives a higher weight than the second sequence-activity relationship.

150. (New) The method of claim 1, wherein the biopolymer of interest is a polypeptide, a polynucleotide, a small inhibitory RNA molecule (siRNA), or a polyketide.

151. (New) The method of claim 1, wherein the biopolymer of interest is a protein kinase, a protein phosphatase, a protease, a receptor, a G-protein coupled receptor, a cytokine, a growth factor or an antigen from an infectious pathogen.

152. (New) The method of claim 1, wherein the biopolymer of interest is a cytochrome P450, a lipase, an esterase, a peptidase, a transferase, a polymerase, or a depolymerase.

153. (New) The method of claim 1, wherein the plurality of positions comprises five or more positions.

154. (New) The method of claim 1, wherein the plurality of positions comprises ten or more positions.

155. (New) The method of claim 119, wherein the plurality of rules comprises five or more rules.

156. (New) The method of claim 119, wherein the identifying combines a score from each rule in a plurality of rules for each respective substitution at each position in the plurality of positions.

157. (New) The method of claim 156, wherein the combining comprises adding (i) a first score from a first rule in the plurality rules and (ii) a second score from a second rule in the plurality rules for the variant of a biopolymer of interest.

158. (New) The method of claim 156, wherein the combining comprises multiplying (i) a first score from a first rule in the plurality rules and (ii) a second score from a second rule in the plurality rules for the variant of a biopolymer of interest.

159. (New) The method of claim 1, wherein the selecting the variant set step b) comprises applying a monte carlo algorithm, a genetic algorithm, or a combination thereof, to construct the variant set, with the provisos that:

(i) each variant in all or portion of the variant set has a number of substitutions that is between a first value and a second value; and

(ii) a number of different pairs of substitutions collectively represented by the variant set is above a predetermined number.

160. (New) The method of claim 159, wherein the first value is two substitutions and the second value is twenty substitutions.

161. (New) The method of claim 159, wherein the first value is four substitutions and the second value is ten substitutions.

162. (New) The method of claim 159, wherein the predetermined number is one hundred.

163. (New) The method of claim 1 wherein

the measuring step c) comprises synthesizing all or the portion of the variants in the variant set, and wherein

the property of a variant in the variant set is an antigenicity of the variant, an immunogenicity of the variant, an immunomodulatory activity of the variant, a catalysis of a chemical reaction by the variant, a thermostability of the variant, a level of expression of the variant in a host cell, a susceptibility of the variant to a post-translational modification, a killing of pathogenic organisms or viruses resulting from activity of the variant or a modulation of a signaling pathway by the variant.

164. (New) The method of claim 1, wherein the sequence-activity relationship has the form:

$$Y = f(w_1x_1, w_2x_2, \ldots w_ix_i)$$

wherein,

Y is a quantitative measure of the property;

x_i is a descriptor of a substitution, a combination of substitutions, or a component of one or more substitutions, at one or more positions in the plurality of positions;

w_i is a weight applied to descriptor x_i; and

f( ) is a mathematical function.

165. (New) The method of claim 164, wherein the modeling comprises regressing:

$$Y = f(w_1x_1, w_2x_2, \ldots w_ix_i).$$

166. (New) The method of claim 165, wherein regressing comprises linear regression, non-linear regression, logistic regressing, or partial least squares projection to latent variables.

167. (New) A plurality of nucleic acid sequences comprising nucleotide sequences encoding all or a portion of the variants in the new variant set of step e) of claim 117.

168. (New) All or a portion of the variants in the new variant set of step e) of claim 117.

169. (New) A population of cells comprising nucleic acid sequences encoding a plurality of variants in the new variant set of step e) of claim 117.

170. (New) The method of claim 159, wherein the predetermined number is thirty.

171. (New) The method of claim 1, wherein each variant in the first plurality of variants is selected on a predetermined basis.

172. (New) The method of claim 1, wherein the value quantifying the confidence with which a substitution in the one or more substitutions of a position in the one or more positions of the biopolymer of interest contributes to the measured property is determined by the method of:

(i) calculating a plurality of sequence activity relationships, wherein each sequence activity relationship in the plurality of sequence activity relationships is calculated using the measured property of an independent subset of the variant set;

(ii) calculating, for each sequence activity relationship in said plurality of sequence activity relationships, a value for the contribution to the measured property by the substitution in the position; and

(iii) calculating a confidence for the value for the contribution to the measured property by the substitution in the position using each said value computed in said calculating step (ii).

173. (New) The method of claim 1 implemented on a computer.

174. (New) A computer program product encoding instructions for implementing the method according to claims 1.